



The Reference Human Genome at NCBI

The Human Genome Project, a 13-year international collaborative effort, reached a major milestone in April 2003 with the release of the first reference sequence for the human genome. This finished sequence follows the working draft, completed in 2001, and described in the February 15 edition of *Nature*.¹ Over the period of the genome sequencing effort, sequencing centers from around the world deposited billions of letters of human DNA sequence into GenBank[®] and its collaborating databases, DDBJ and EMBL, where the data was immediately made available to researchers.

Between the appearance of the draft sequence and the release of the finished reference sequence, NCBI maintained interim assemblies of the data to ensure access by the research community to the most complete genome draft. Now, the finished reference human genome sequence, along with the results of NCBI analysis and annotations, is available for viewing and downloading.

The Genome and its Genes

The reference human genome, a small portion of which is shown in Figure 1, consists of 24 finished chromosomes of 2.9 billion bases and covers about 99 percent of the gene-containing DNA. The sequence is accurate, on average, to the level of one error per 10,000 bases. Small updates to the assembly will continue as complex regions are further refined and the small number of remaining gaps between the large stretches of contiguous sequence, or “contigs”, are closed.

NCBI identifies known genes in the genome by aligning Reference Sequence (see RefSeq below) and GenBank mRNAs to the assembled

continued on page 3

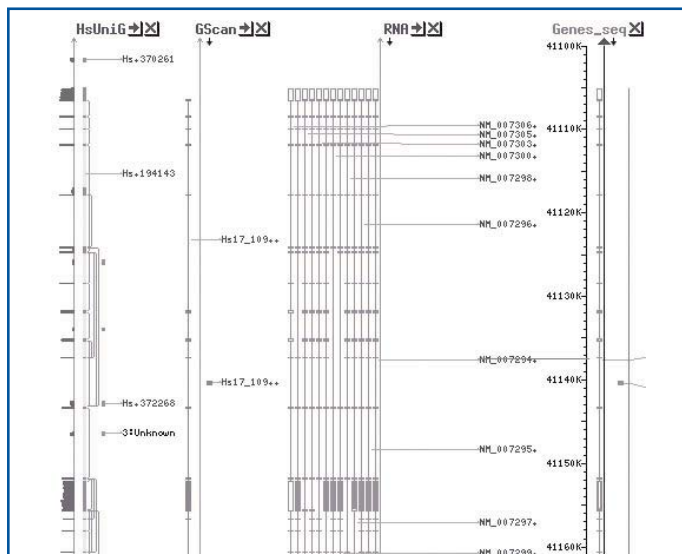


Figure 1. Map Viewer display for the human BRCA1 gene showing, from the right, the NCBI gene model, 13 transcript variants, a GenomeScan predicted gene model, and UniGene cluster sequences that map to the region.

SARS Coronavirus Resource

The first complete sequence of the SARS Coronavirus, determined by the BC Cancer Agency Genome Sciences Centre in Canada, was submitted to GenBank prior to publication as an unannotated nucleotide sequence and assigned GenBank accession number AY274119. The sequence was subsequently processed through the NCBI viral genome annotation pipeline and made available in Entrez Genomes under RefSeq accession NC_004718 as the *SARS-CoV* reference sequence within about 24 hours of its submission. The results of this computational analysis can be accessed from the

continued on page 4

In this issue

- 1 The Reference Human Genome
- 1 SARS Coronavirus Resource
- 2 Gene Expression Omnibus (GEO)
- 5 Major Histocompatibility Complex database (dbMHC)
- 5 RefSeq Release 1 Ready for Download
- 5 GenBank Release 137
- 6 New Microbial Genomes in GenBank
- 6 Sequence Revision History Page Offers New Comparison Function
- 7 BLAST Lab

NCBI News is distributed four times a year. We welcome communication from users of NCBI databases and software and invite suggestions for articles in future issues. Send correspondence to *NCBI News* at the address below.

NCBI News
National Library of Medicine
Bldg. 38A, Room 3S-308
8600 Rockville Pike
Bethesda, MD 20894
Phone: (301) 496-2475
Fax: (301) 480-9241
E-mail: info@ncbi.nlm.nih.gov

Editors

Dennis Benson
David Wheeler

Contributors

Tanya Barrett
Susan Dombrowski
Wolfgang Helmberg
Tao Tao

Writers

Vyvy Pham
David Wheeler

Editing and Production

Robert Yates

Graphic Design

Robert Yates

In 1988, Congress established the National Center for Biotechnology Information as part of the National Library of Medicine; its charge is to create information systems for molecular biology and genetics data and perform research in computational molecular biology.

The contents of this newsletter may be reprinted without permission. The mention of trade names, commercial products, or organizations does not imply endorsement by NCBI, NIH, or the U.S. Government.

NIH Publication No. 03-3272

ISSN 1060-8788

ISSN 1098-8408 (Online Version)

New Data Query and Visualization Tools for Gene Expression Omnibus (GEO)

The Gene Expression Omnibus (GEO) database, the first public repository for gene expression data, premiered at NCBI in July 2000. The GEO database contains a wide assortment of high-throughput experimental data, including single and dual channel microarray-based experiments measuring the abundance of mRNA, genomic DNA and protein molecules. Data from non-array-based high-throughput functional genomics and proteomics technologies are also archived, including serial analysis of gene expression (SAGE) and protein identification technology. To date, the GEO database contains data representing almost 10,000 hybridization experiments and SAGE libraries from 30 different organisms.

Several new tools and features have been developed to enable effective exploration, visualization and analysis of the data in GEO. To create these tools, GEO data are first assembled into comparable sets, or GEO DataSets (GDS). A GDS represents a collection of biologically and statistically comparable GEO samples. Two new databases have been created to query these datasets - Entrez GEO and Entrez GEO DataSets. Entrez GDS queries dataset definitions and original experimental annotation to facilitate identification of experiments of interest. Entrez GEO displays individual gene expression/molecular abundance profiles from each dataset.

Searching and browsing GEO

Several methods are available for searching, browsing and retrieving data from GEO. Specific GEO records may be retrieved by entering a valid GEO accession number into the Accession Display toolbar on the GEO Home Page. A listing of the current holdings in GEO is accessi-

ble from the 'Repository Browser' link on the GEO home page. The 'DataSet Browser' link displays the full collection of GDS's, which can be sorted alphabetically by title, platform type, GEO platform (GPL) identifier, organism, and GDS accession. Sophisticated searches of GEO data and linking to other Entrez databases can be accomplished using Entrez GEO and Entrez GDS. The Quick Query Builder available on the GEO home page facilitates popular Entrez GEO/Entrez GDS query construction.

To search for an experiment of interest, submit a query under the 'Datasets' tab, or from the GEO DataSet database in Entrez. This initiates a search of all dataset annotation including the GDS description, reference series and sample descriptions, titles, keywords, source material, contributor, authors and organisms, as well as some general technical information including experiment type, probe type and value measurement type. The results will list all datasets that fit the user-defined search criteria.

To search for individual gene expression/molecular abundance profiles of interest, submit a query under the 'Gene Profiles' tab, or from the GEO database in Entrez. A particular gene or molecule of interest may be searched for by gene name, symbol or alias, or using sequence identifiers such as GenBank accession numbers, clone IDs or ORF names. Several parameters are available to refine an Entrez GEO search and help identify interesting or significant molecular abundance profiles. GEO datasets are partitioned into subsets which reflect experimental design. Queries can be made for differences related to a specific experimental variable such as age, developmental

continued on page 8

Reference Human Genome continued from page 1

genomic sequence using MegaBLAST. NCBI also predicts genes computationally, but includes predicted genes in the annotation only if they do not overlap a gene model based on an mRNA alignment. About 25,000 genes have been annotated on the genome using these two methods. Sequence variations are mapped to the reference genome via BLAST[®], using the data in the Database of Single Nucleotide Polymorphisms (dbSNP). For more information on NCBI's human genome assembly or annotation, see the Web pages referenced in the box entitled "Human Genome Build Information".

Exploring the Genome

The Human Genome Resource page, found under 'Hot Spots' on the NCBI Home Page, provides an entry into the NCBI resources, databases, and tools related to the human reference genome. Three primary resources accessible from this page, as well as from the NCBI Home Page, are RefSeq, LocusLink, and the Map Viewer.

RefSeq and LocusLink

The human portion of the RefSeq database (for more information, see "RefSeq Release 1 is Ready for Download", this issue) includes the transcript and associated protein sequences derived from GenBank submissions, the gene models derived from the genome by prediction, and the contig and chromosomal records for the reference genome itself. RefSeqs are recognized by accession numbers beginning with two letters, indicating the type of sequence, and an underscore. Transcript and protein RefSeqs with the prefixes "NM_" and "NP_", respectively, are derived from GenBank submissions and therefore are considered to be experimentally

supported to some degree. Predicted transcripts and their protein translation products bear, respectively, the prefixes "XM_" and "XP_".

Genomic contigs begin with "NT_" while reference records for the 24 human chromosomes comprise the series "NC_000001-NC_000024". The RefSeq contigs, transcripts, and proteins are also retrievable with standard Entrez queries by accession number, gene symbol, or protein name and can be restricted to the RefSeq entries using 'Entrez Limits'.

LocusLink offers a single query interface to gene loci for many organisms, and includes all human genes defined by the genome annotation process. LocusLink reports display descriptive information and links to related NCBI resources such as RefSeq, NCBI's Map Viewer,

or symbols, marker names, SNP identifiers, accession numbers and other identifiers makes it easy to navigate to a gene or region of interest. The Map Viewer for the human reference genome displays cytogenetic maps, physical maps, maps showing predicted gene models, EST alignments with links to UniGene clusters from human and related organisms, and mRNA alignments used to construct gene models. A tabular view of the data allows convenient export of the information shown in the graphical display. Map Viewer displays are linked to supporting resources such as LocusLink, the Evidence Viewer, and Model Maker; the latter two tools are described in the shaded box entitled "Human Genome Tools". Segments of the genomic assembly shown in the

Human Genome Tools

Model Maker (MM)—allows the construction of transcript models using the pre-computed alignments of NCBI RefSeqs, GenBank transcripts, ESTs, and predicted transcripts to the NCBI human genome assembly.

The Evidence Viewer (EV)—displays the alignments of RefSeq transcripts, GenBank mRNAs or predicted transcripts, and ESTs that support an NCBI gene model. The EV produces a graphical summary along with a detailed exon-by-exon view and shows any proteins annotated on the transcript.

Human Genome BLAST—The human genome BLAST page offers MegaBLAST for rapid searches of the reference genome. Standard variants of BLAST are also available to search the RefSeq transcripts and proteins arising from NCBI annotations. Human Genome BLAST hits are displayed in the MapViewer to show their genomic context.

Evidence Viewer, Model Maker, BLAST Link, UniGene, protein domains from NCBI's Conserved Domain Database, and the Homologene database. Follow the links under 'Hot Spots' on the NCBI Home Page to reach the LocusLink and RefSeq pages.

The Map Viewer

The NCBI Map Viewer, available under 'Hot Spots' on the NCBI Home Page and via the Entrez Links menu for nucleotide and protein records shown in the Map Viewer, generates graphical views, such as that shown in Figure 1, of aligned chromosomal maps for human and other organisms. A flexible query interface that supports gene names

graphical view may be downloaded using the Map Viewer's "Download/View Sequence" link. A Map Viewer help document is available via the "Human Maps Help" link on the Map Viewer page. See also chapters in the NCBI Handbook, available by clicking "NCBI Handbook" under 'Hot Spots' on the NCBI Home Page.

—DW

¹Nature. 2001 Feb 15;409(6822):745-964.

Human Genome Build Information

(see "Available Documentation" links at the bottom of the Web page)

www.ncbi.nlm.nih.gov/mapview/map_search.cgi

SARS Coronavirus Resource
continued from page 1

Analysis section of the new SARS Coronavirus Resource, a Web page providing a point of access to sequence data and a wealth of other information about the SARS Coronavirus. The types of analyses available on the SARS Coronavirus Resource page are described below.

Pair-wise global alignments of NC_004718 with other viral genomic sequences, pre-computed using the “band” version of the Needleman-Wunsch algorithm, are shown in graphical representations highlighting mutations, deletions, and insertions among the sequences as shown in Figure 1. Global alignments are updated automatically as new virus sequences enter GenBank.

Predicted SARS proteins are listed in a separate table, complete with information on the corresponding gene, accession number, length, and pre-computed comparison to other proteins. Pre-computed alignments of *SARS-CoV* protein sequences from the RefSeq collection of complete genomes in Entrez, are accessible from the column “mA” in the table. The alignments, such as that shown in Figure 2, were constructed using the ClustalX program and, in some cases, manually edited. Similarities are highlighted in color, if at least 80% of residues in a column are identical or fall into at least one of the following amino acid groups: aromatic (FHWY), aliphatic (ILVA), hydrophobic (ACFILMVWY), alcohol (STC), charged (DEHKR), polar (CDEHKNQRST), tiny (AGS), small (ACDGNPSTV), or bulky (EFIKLMQRWY).

SARS protein sequences, compared by BLAST/PSI-BLAST to sequences with known 3D structures, are listed in the Related Structures section. The links present sequence alignments to 3D structures and mapping displays using the Cn3D molecular graphics viewer. Additional related

structures were selected from the VAST 3D-structure neighbors of the proteins identified by BLAST/PSI-BLAST. Structure links are also updated automatically as new data enters the databases.

In addition to the sequence analysis performed, automatic searches of SARS-related information in the Entrez databases—PubMed, Genomes, Nucleotide, Protein, Structures—are provided. Links to resources, such as the Center for Disease Control and the World Health Organization, are listed to provide comprehensive disease information. Access the SARS resource from Entrez Genome or directly at:

www.ncbi.nlm.nih.gov/genomes/SARS/SARS.html

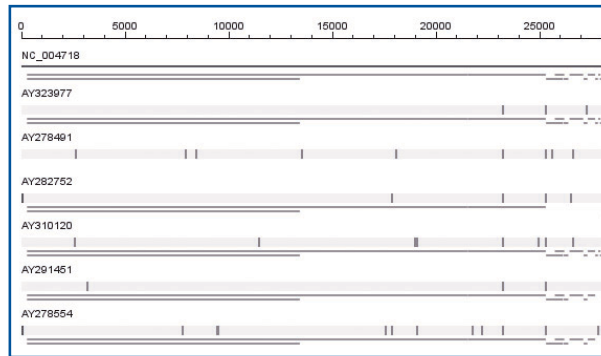


Figure 1: Graphical format of global alignment of NC_004718 with other viral genome sequences. The alignment is updated automatically as new virus sequences enter GenBank.

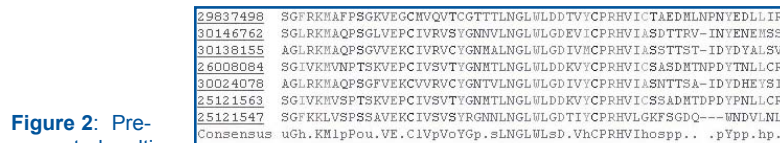


Figure 2: Pre-computed multiple alignment of the coronavirus nsp2 proteins.

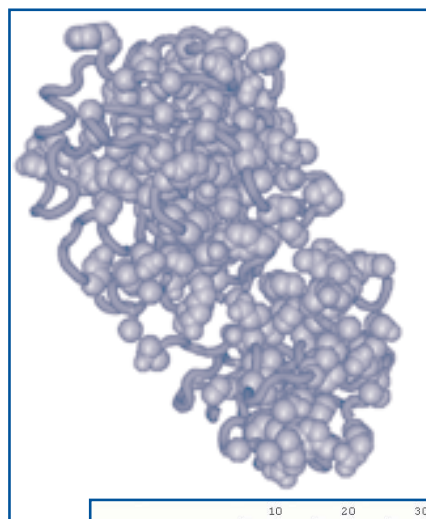
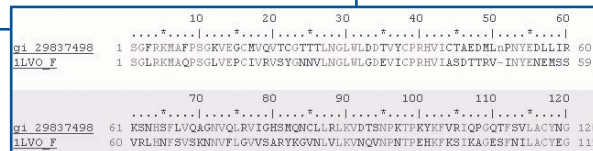


Figure 3a and 3b: Portion of the alignment of the sequence of putative coronavirus nsp2 protein NP_828863 to that of the coronavirus main proteinase from Transmissible gastroenteritis virus, Protein Databank code 1LVO, Chain F. In the sequence alignment, identical residues are darker. In the Cn3D rendering of the structure of 1LVO, residues that are identical in the alignment are shown in a spacefilling representation while intervening residues are shown as a backbone trace, illustrating a strong degree of sequence conservation in core regions of the protein.



Major Histocompatibility Complex database (dbMHC)

The dbMHC database provides an open, publicly accessible platform for DNA, and clinical data related to the human Major Histocompatibility Complex (MHC). The need to share research and clinical data focused on the MHC has led to a series of meetings at the International HLA WorkShop & Congress (IHWC). The data generated from the 13th IHWC is presented at NCBI in dbMHC. In addition, the dbMHC will provide tools for submission and analysis of research data linked to the MHC. Users can access dbMHC at:

www.ncbi.nlm.nih.gov/mhc

The dbMHC is divided into two main sections, a Reagent Database section and a Clinical section. The Reagent database contains the reagent data needed to trace DNA typing. This section provides an open platform for the submission, evaluation, and editing of individual reagent specifications of Sequence Specific Oligonucleotides and Sequence Specific Primers as well as typing kit information. All reagents are characterized for allele specificity

using the current curated World Health Organization HLA allele database in cooperation with IMGT/HLA.

The dbMHC offers several resources for the analysis and display of the MHC and KIR region, e.g. an interactive formatting sequence retrieval tool, and a Sequencing Based Typing tool, capable of aligning and interpreting heterozygote sequences. The database resource also features dbMHCms, a tool to search descrip-

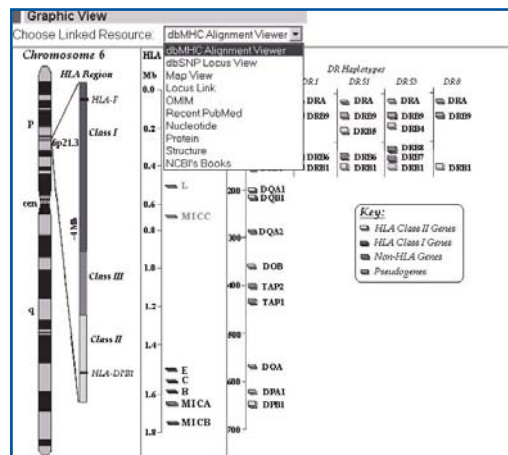


Figure 1. Graphical view of the human MHC region on chromosome 6. MHC genes are marked and are linked to one of the related resource selected using the pull-down menu.

tive information for known short tandem repeats within the MHC.

The Clinical section will contain anonymous clinical data from individuals taking part in MHC-related research projects in the general categories of Anthropology, Cytokine Polymorphisms, HLA-E,F,G, Cancer, Disease, HLA Alloantibodies & Kidney Graft Rejection, Mycobacterial Disease, New Allele Registry, Hemochromatosis/Psoriasis, and Virtual DNA Analysis. The data for these projects will be made available in the near future.

A graphical view of the MHC region of human chromosome 6, shown in Figure 1, is also provided. The graphic highlights the MHC genes in the region and provides links to related resources for each gene. The particular resource to which the graphic is linked is selected using a pull-down menu. Available resources include the Map Viewer, dbSNP, Entrez nucleotides and Entrez proteins and others, however, the default resource is the new dbMHC Alignment Viewer that displays alignments between MHC genes belonging to various haplotypes.

—WH

RefSeq Release 1 is Ready for Download

The Reference Sequence (RefSeq) project aims to provide a non-redundant set of genomic, transcript, and protein sequences, for a wide spectrum of organisms. RefSeqs provide a stable reference for gene identification and characterization, mutation analysis, expression studies, polymorphism discovery, and comparative analyses and are used in the functional annotation of genomes, including those of mouse and human. The first release of a complete NCBI RefSeq database is now ready for download at: ftp.ncbi.nih.gov/refseq/release/

The release includes over 785,000 protein sequences and more than 200,000 genomic, and mRNA, sequences for about 2,000 organisms. Available FTP files include the RefSeq sequences, a catalog of the release contents, statistics, and documentation. Subsets for taxonomic genes are available in subdirectories for “invertebrates”, “plants”, “mitochondria”, “vertebrates-mammalian”, “microbial” and several others where the RefSeq records are offered in GenBank, GenPept, and FASTA formats.

The records for the entire database are found in the “complete” subdirectory and are offered in binary ASN.1 format in the case of nucleotide and protein records, and, additionally, in FASTA format for protein records. To subscribe to the NCBI's refseq-announce mailing list, and receive announcements of future RefSeq releases, or to read more about the RefSeq project, visit the RefSeq Home Page at:

www.ncbi.nlm.nih.gov/RefSeq/

GenBank Release 137

GenBank release 137 (August 2003) contains over 27 million sequence entries totaling more than 33 billion base pairs. GenBank is accessible via the Entrez search and retrieval system. The flatfile and ASN.1 versions of the release are found in the “genbank” and “ncbi-asn1” directories respectively at:

ftp.ncbi.nih.gov/

Uncompressed, the release 137 flatfiles consume about 121 gigabytes while the ASN.1 version consumes about 100 gigabytes. The data can also be downloaded at two mirror sites:

genbank.sdsc.edu/pub

bio-mirror.net/biomirror/genbank

New Microbial Genomes in GenBank

Organism	GenBank RefSeq Accession Numbers	Organism	GenBank RefSeq Accession Numbers
<i>Bordetella bronchiseptica</i> str. RB50	BX470250 NC_002927	<i>Streptomyces avermitilis</i> MA-4680	chromosome: BA000030 NC_003155 plasmid: AP005645 NC_004719
<i>Bordetella parapertussis</i> str. 12822	BX470249 NC_002928	<i>Chlamydomophila caviae</i> GPIC	chromosome: AE015925 NC_003361 plasmid: AE015926 NC_004720
<i>Bordetella pertussis</i> str. Tohama I	BX470248 NC_002929	<i>Candidatus blochmannia floridanus</i>	BX248583 NC_005061
<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> str. CCMP 1375	AE017126 NC_005042	<i>Haemophilus ducreyi</i> 35000 HP	AE017143 NC_002940
<i>Chlamydomophila pneumoniae</i> TW-183	AE009440 NC_005043	<i>Helicobacter hepaticus</i> ATCC 51449	AE017125 NC_004917
<i>Pirellula</i> sp.	BX119912 NC_005027	<i>Streptomyces avermitilis</i> MA-4680	chromosome: AE016877 NC_004722 plasmid: AE016878 NC_004721
<i>Mycoplasma gallisepticum</i> R	AE015450 NC_004829	<i>Coxiella burnetii</i> RSA 493	AE016828 NC_002971
<i>Shigella flexneri</i> 2a str. 2457T	AE014073 NC_004741	<i>Synechococcus</i> sp. WH 8102	BX548020 NC_005070
<i>Nitrosomonas europaea</i> ATCC 19718	AL954747 NC_004757	<i>Rochlorococcus marinus</i> str. MIT 9313	BX548175 NC_005071
<i>Bacillus anthracis</i> str. Ames	AE016879 NC_003997	<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> str. CCMP1378	BX548174 NC_005072
SARS Coronavirus Tor2	AY274119 NC_004718		

For more detailed information, see the online version of the Summer 2003 NCBI News, or use the GenBank or RefSeq Accession Number to query Entrez "Genome" database using the query box on the NCBI Home Page.

Sequence Revision History Page Offers New Comparison Functions

Changes in records from the Entrez Nucleotide and Protein databases can now be visualized using the Sequence Revision History tool at:

www.ncbi.nlm.nih.gov/entrez/sutils/girevhist.cgi

To retrieve the history of a record, one can enter the accession number, the GI number, or the FASTA-style SeqID of that record into the query box at the top of the page. Figure 1 shows the revision history table for the nucleotide sequence AF123456. The table displays the dates of changes made to the record since it was first released. The first two columns contain, respectively, the GI numbers and the version numbers of the sequence, which are only changed if the sequence data itself has been modified. Updates to any record may include changes to the sequence, publication information, or annotations made by the authors. In this case, the sequence data has been modified once. The "Update Date", identical to the "Modification Date" seen on the Entrez Limits page, is the date that appears in the upper right hand corner

in the flat file view of Entrez records. The "Status" column indicates the most recently-updated version of the record, the "Live" record, that is retrievable from Entrez; the older versions have a status of "Dead". The date the record was first seen at NCBI is given at the bottom of the table.

The last two columns are labeled I and II, respectively, and are used to select the two versions of the record to compare. One can view the differ-

ences in several display formats, including GenBank/GenPept flat file, XML, and ASN.1. The differences between the records are highlighted in color, as shown in Figure 2. The "GenBank diff" format lists the differences in the manner of the UNIX "diff" command, without highlighting. The FASTA and BLAST formats can be used to pinpoint differences in the sequence data.

—VP

Figure 1. The Sequence Revision History table for AF123456. The last two versions of the record have been selected for comparison.

GI	Version	Update Date	Status	I	II
6633795	2	Jul 25 2000 8:09	Live	<input checked="" type="radio"/>	<input type="radio"/>
6633795	2	Jan 28 2000 5:18	Dead	<input type="radio"/>	<input checked="" type="radio"/>
6633795	2	Dec 23 1999 2:25	Dead	<input type="radio"/>	<input type="radio"/>
4454562	1	Mar 23 1999 1:24	Dead	<input type="radio"/>	<input type="radio"/>
4454562	1	Mar 20 1999 12:06	Dead	<input type="radio"/>	<input type="radio"/>

Accession AF123456 was first seen at NCBI on Mar 20 1999 12:06

GI	GenBank/GenPept	Update Date
6633795	GenBank/GenPept	Jul 25 2000 8:09
6633795	GenBank Diff	Jan 28 2000 5:18
	ASN.1	
	FASTA	
	XML	
	BLAST	

LOCUS AF123456 1510 bp mRNA linear VRT 25-JUL-2000
 LOCUS AF123456 1510 bp mRNA linear VRT 28-JAN-2000
 DEFINITION Gallus gallus doublesex and mab-3 related transcription factor 1 (DMRT1) mRNA, partial cds.
 ACCESSION AF123456
 VERSION AF123456.2 GI:6633795
 KEYWORDS .
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

Figure 2. GenBank display of the differences between two versions of AF123456. Sections of the record in which there are differences are shown for each version and highlighted.

Gene Expression Omnibus *continued from page 2*

stage, disease state, and others. The relative abundance of a particular molecule and the degree of measurement variability are also search parameters in GEO. Results are returned as a set of pre-computed molecular abundance profiles such as that shown in Figure 1 for human synaptopodin 2. Results are returned in order of most-interesting-first, based on a scoring scheme which considers statistically significant differences, expression level, outliers, and variability.

Entrez GEO may also be queried for sequences of interest based on nucleotide sequence similarity, thus facilitating the identification of sequence homologs of interest, e.g., related gene family members or for cross-species comparisons across all GEO datasets. The “Sequence BLAST” search function on the

GEO Home Page accepts either a FASTA sequence, GI number or accession number as input and performs a BLAST search against all the sequences represented on microarray platforms or SAGE libraries in GEO.

Within Entrez GEO results, following the “Profile Neighbors” link from selected expression profiles will display those probes within the same dataset that show an expression profile that is similar to the one selected.

When the “Sequence Neighbors” link is selected, the results will be those sequences that are similar or identical to the query probe over all GEO datasets. Entrez GEO and Entrez GDS retrieval results are fully integrated with each other as well as other Entrez databases including Nucleotide, UniGene, MapViewer and PubMed.

All original GEO records as well as GDS data are available for download at:

<ftp.ncbi.nih.gov/pub/geo/data/>

Questions regarding the submission of data to GEO may be sent to:

geo@ncbi.nlm.nih.gov

General inquiries about GEO may be sent to the geo alias or to the NCBI Help Desk:

info@ncbi.nlm.nih.gov

—TB, SD

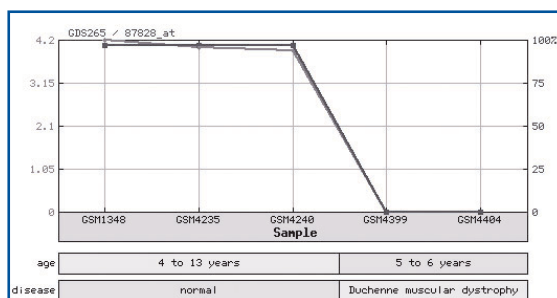


Figure 1. Expression profile for human synaptopodin 2 from GDS265 showing decreased expression in tissues taken from muscle biopsies of Duchenne muscular dystrophy patients relative to controls.

Department of Health and Human Services
Public Health Service, national institutes of Health
National Library of Medicine
National Center for Biotechnology Information
Bldg. 38A, Room 3S308
8600 Rockville Pike
Bethesda, Maryland 20894

Official Business
Penalty for Private Use \$300

FIRST CLASS MAIL
POSTAGE & FEES PAID
PHS/NIH/NLM
BETHESDA, MD
PERMIT NO. G-816